

Rubix ML

Please help test the code: [Machine Learning](#) and provide use cases for [Machine Learning for Email](#), [Machine Learning for lead scoring](#) and [Machine Learning for SEO and SEM](#). Also usable for [Spellcheck](#).

What

Integrate <https://github.com/RubixML/RubixML> in Tiki. Rubix ML is a "high-level machine learning and deep learning library"

What is machine learning? Please see: <https://docs.rubixml.com/2.0/what-is-machine-learning.html>

Both Tiki and Rubix ML are written in PHP, which will facilitate the integration. This is major: Most of the alternatives to Rubix ML are written in Java or Python. We could use them but since they would not be built-in, only a tiny fraction of the Tiki community would have access.

Tiki already has mature data management tools

. Now, in close collaboration with the Rubix ML community, we will add the necessary tools and interfaces to become a complete machine learning platform (managing data, choosing a model, training, evaluation, etc.) accessible to power users, like the rest of Tiki. We will contribute to Rubix ML and make it easier for all other PHP Open Source projects to also integrate with Rubix ML.

Where

- We are coordinating on the Rubix ML chat room, powered by Telegram: <https://t.me/RubixML>
 - Telegram client apps are Open Source: <https://telegram.org/>

Why

- Permits multiple new features.
 - See "Related" section below for some examples. Many of these features have been desired for years but we didn't have a clean solution. Both Rubix ML and Tiki have a large feature set and a "one stop shop" philosophy.
 - See various sections at <https://github.com/RubixML/RubixML/tree/master/src> like Anomaly Detection, Classification, Clustering, etc.

On what types of data?

- On Tiki system data (ex.: logins logs) so will be providing insight for all Tiki instances!
 - Spammy registrations
- On standard features like forums, wiki page, comments, email, etc.
 - Email classification, Spam detection
- On ad hoc data structures made with <https://tikitrackers.org/>

Who

- Marc (instigator)
- Andrew (Lead dev of Rubix ML) is providing guidance)
- Roberto (developer)
- Victor (Back-end code) ~~will do~~ did the initial integration
- Jonny (Front-end code)
- [Alain Désilets](#) (advisor)
- Ricardo Melo (advisor)

- Michael I. (tester/requirements for a multilingual project)

How

We'll start with some simple use cases, like reproducing some of the "Project Spotlight" on <https://rubixml.com/> [↗](#), but directly within Tiki.

About performance

Performance is very important to train the model. Here is Andrew DalPino, the founder of Rubix ML:

Slides:

https://docs.google.com/presentation/d/1a08XvUzA_9RHtBf5S-FOv1XBLgMN7u9dY-Z_EI3VXPE/edit?usp=sharing [↗](#)

For developers

If you are new to Rubix ML

- Read all the [documentation](#) [↗](#)
- Do a quick review of all the code base
- Run at least one of the tutorials: <https://docs.rubixml.com/en/latest/#tutorials-example-projects> [↗](#)
- Read [all the open issues](#) [↗](#)
- Contribute <https://github.com/RubixML/RubixML/blob/master/CONTRIBUTING.md> [↗](#)

If you are new to Tiki

Reading all the documentation and even a quick scan of all the source code is an unrealistic goal because the project is huge. So just focus at first on Tiki Trackers

1. Join
 - <https://gitter.im/tiki-org/community> [↗](#) (where Tiki devs hang out)
 - <https://tiki.org/login> [↗](#) (documentation, forum, bug reports, etc.)
 - [Dev Mailing List](#) (where decisions are taken)
2. Read all the content and watch all the videos at <https://tikitrackers.org/> [↗](#)
3. [Install Tiki](#)
 - You can get source from <https://gitlab.com/tikiwiki/tiki> [↗](#) instead of tarball/zip
4. Explore Tiki features for a few hours
5. Build a simple tracker for yourself
6. Contribute code to Tiki: [Git Workflow](#)

Once you know both Rubix ML and Tiki Trackers

- Think about how we could add a graphical user interface (GUI) to Tiki to leverage of Rubix ML.
- Think about how the Rubix ML demos could be handled within Tiki
- Think about how we can have something like MLT without Elasticsearch:
<https://github.com/RubixML/RubixML/issues/75> [↗](#)


Other Ideas For Rubix ML in Tiki

Actually more like questions - could Rubix ML do things like this?

- **Duplicate content prevention?**
(to display a warning when posting something covered by other content, e.g. in Forums or Trackers)
- **Automatic forum moderation** - more than just avoiding prevented vocabulary?
- **Reading/scraping .pdf's**, e.g. resumes? First, each person formats their resume differently, second,

there can be different terminology (e.g. 'work experience' 'employment history'). Maybe in combination with OCR?

- **Mobile receipt capture application**

- similar to: "Sensibill Capture is a mobile receipt capture application. The Sensibill Receipt Data Extraction API Track this API uses machine learning and Optical Character Recognition (OCR) to extract pertinent data from receipts and other documents. The API enables users to extract over 150 data points, including merchant, total, items, taxes, and more from a given receipt. Developers need to contact the provider for API access and documentation. This API is listed in the OCR category of the ProgrammableWeb API directory." (source: programmableweb.com )

- **Reading Merchandiser Quotation** *similar to the above, 'Mobile Receipt Capture Application'*

- A merchandisers receive quotations for products that have product names, and/or product codes and quantities in them. The quotations are in the form of plain email texts, word documents or spreadsheets. Each quotation has explanatory labels and descriptive sentences but each one has its own way of presenting the information. This would traditionally be solved by a person manually going through the sources and adding the info in a database. ML would need to open and read these various formats than populate the database.

- **Farming Automation:** When deploying ground bots for mechanical weed elimination, use ML to differentiate all weeds from the particular crop being grown. Example video from the University of Illinois program:

- **Farming Automation** - a very advanced use case (but it's good to set lofty goals): while a group of drones (100+) are completing an assignment to spray crop protection products (herbicides, pesticides, etc.) on a specific field, several of them will fail. RubixML should recognize when they fail, send out other drones to pick them up, re-organize the workload to account for the sudden changes and send out replacement drones (all controlled by Trackers / List Execute / Scheduler / ML)

- ...

Stats

Projects

With Drones

<https://www.sciencedirect.com/science/article/pii/S2095809918308130> 

Related

- [https://www.zdnet.com/google-amp/article/six-applications-for-deep-learning-as-it-enters-the-enterpris
e-mainstream/](https://www.zdnet.com/google-amp/article/six-applications-for-deep-learning-as-it-enters-the-enterprise-mainstream/) 
- <https://www.linkedin.com/groups/8952251/> 
- [Machine Learning](#)
- [Naive Bayes classifier](#)
- [Natural language processing](#)
- [Auto-Classification](#)
- [Optical character recognition](#)
- [Text Mining](#)
- [IRC QA Mining](#)
- [Natural Language Generation](#)
- [Farming Automation](#)
- [Use Cases for NLP and IR in Tiki](#)
- <https://developer.ibm.com/exchanges/data/> 
- [Follow up about "Is PHP Now Suitable For Machine Learning?"](#) 
- [Widows and Orphans in mPDF](#) (We will attempt to solve with machine learning)

Machine Learning vs Artificial Intelligence

In French

Video in French about Machine Learning in PHP

Slides:

<https://docs.google.com/presentation/d/1XuxgQtIcXuSnLRxSJNPaZBjEPyzg8tj-RzE4wOykyDg/edit?usp=sharing> 

Other examples

- NextCloud email classification:

<https://github.com/nextcloud/mail/blob/master/lib/Service/Classification/ImportanceClassifier.php> 